In J. T. Wixted (Ed.), The Stevens' handbook of experimental psychology and cognitive neuroscience, Volume 4 (4th ed., pp. 395-427). New York, NY: Wiley.

CHAPTER 12

Implicit Social Cognition

ADAM HAHN AND BERTRAM GAWRONSKI

INTRODUCTION

The term "implicit social cognition" is conventionally used to refer to research in social psychology that uses a particular class of computerized measurement instruments to infer thoughts and affective reactions without directly asking participants to report on them. A central feature of these instruments is that they limit participants' ability to strategically control their responses, which distinguishes them from traditional instruments that rely on self-report (Gawronski & De Houwer, 2014). The measurement outcomes of these computerized instruments are commonly referred to as implicit measures, and the measurement outcomes of traditional self-report instruments are usually called explicit measures.

A common way to conceptualize the constructs of implicit social cognition refers to the idea of mental association, most notably evaluative and semantic associations (Greenwald et al., 2002). For example, the construct of attitude can be defined as the mental association between an attitude object and a positive or negative evaluation (e.g., association between *pizza* and *good*). Moreover, whereas the term "prejudice" refers to the mental association between a social group and a particular evaluation (e.g., association between *Muslims* and *bad*), the term "stereotype" can be defined as the mental association between a social group and a semantic attribute (e.g., association between *women* and *warm*). Similarly, the term "self-esteem" refers to the association between the self and a particular evaluation (e.g., association between *self* and *good*), and the *self-concept* refers to associations between the self and semantic attributes (e.g., association between *self* and *extraverted*). A valuable aspect of the concept of mental association is that it can be applied to a wide range of objects that are of interest to psychologists (e.g., consumer products, political parties). Although alternative frameworks have been proposed that reject the idea of mental association (see Hughes, Barnes-Holmes, & De Houwer, 2011), associative theorizing has been a driving force in research on implicit social cognition, including the development of measurement instruments and the generation of empirical predictions. The basic idea is that mental associations can be activated automatically, and this automatic activation in turn influences responses on the measurement instruments of implicit social cognition.

WHAT IS "IMPLICIT" ABOUT IMPLICIT SOCIAL COGNITION?

Although the term "implicit social cognition" was initially interpreted in a broader sense

(Greenwald & Banaji, 1995), it has become a descriptive label for social psychological research that uses the above-mentioned class of measurement instruments (Gawronski & Payne, 2010). However, why exactly the measurement outcomes of these instruments should be described as "implicit" is still a matter of debate. The two most prominent positions in this debate can be traced back to different historical roots of this particular research field (Payne & Gawronski, 2010).

The first line of research was inspired by cognitive research on automatic processes in attention and emerged from the desire to overcome the problems of social desirability in research using self-reports (Fazio, Jackson, Dunton, & Williams, 1995). Using sequential priming tasks (see the subsection titled "Sequential Priming Tasks"), this line of work was primarily concerned with the automatic activation of attitudes, showing that attitudes can influence evaluative responses even when participants do not have the intention to evaluate the attitude object. An important assumption underlying this research is that the impact of automatically activated attitudes on self-reports is reduced when participants are motivated and able to control their responses (Fazio, 2007). Thus, in this line of work, the implicit-explicit distinction is typically used to describe different kinds of measurement instruments, in that implicit measures are conceptualized as instruments that limit the opportunity for strategic control and explicit measures are conceptualized as instruments that permit strategic control.

The second line of research, also concerned with the lack of honest self-reports, grew out of cognitive research on implicit memory (Greenwald & Banaji, 1995). The central assumption underlying this work is that prior experiences can influence responses even when participants are unable to verbally report on those experiences. Based on this idea, Greenwald and Banaji (1995) defined implicit social cognitions as "introspectively unidentified (or inaccurately identified) traces of past experience that mediate responses" (p. 5). Although this definition specified past experiences as the inaccessible component, it has often been misinterpreted as indicating that the mental contents resulting from these experiences are inaccessible to introspection. Thus, in this line of research, the implicit–explicit distinction is typically used to distinguish between mental contents that are conscious versus unconscious. For example, whereas self-reports are assumed to reflect explicit attitudes (i.e., conscious attitudes), the new class of computerized instruments is assumed to capture implicit attitudes (i.e., unconscious attitudes).

To resolve the terminological confusion surrounding the implicit–explicit distinction, De Houwer, Teige-Mocigemba, Spruyt, and Moors (2009) suggested using the terms "implicit" and "explicit" to describe measurement outcomes rather than measurement instruments or psychological constructs. According to this conceptualization, a measurement outcome can be called implicit when the to-be-measured attribute (e.g., attitude, self-concept) influences the observed outcome in an automatic fashion (i.e., when the impact of the attribute on participants' responses is unintentional, unconscious, resource-independent, or uncontrollable; Bargh, 1994). Conversely, a measurement outcome should be called explicit when the to-be-measured attribute influences the observed outcome in a controlled fashion (i.e., when the impact of the attribute on participants' responses is intentional, conscious, resource-dependent, or controllable; Bargh, 1994). Different from the implicit versus explicit nature of measurement outcomes, measurement instruments may be described as direct or indirect, depending

on whether they require a self-assessment of the to-be-measured attribute. According to this conceptualization, a measurement instrument is direct when it is based on participants' self-assessment of the to-be-measured attribute (e.g., when participants' racial attitudes are inferred from their self-reported liking of Blacks versus Whites). Conversely, a measurement instrument is indirect when it is not based on a self-assessment (e.g., when participants' racial attitudes are inferred from their speed in responding to positive and negative words after brief presentations of Black versus White faces) or when the to-be-measured attribute is inferred from a self-assessment of attributes other than the to-be-assessed attribute (e.g., when participants' racial attitudes are inferred from their self-reported liking of neutral objects after brief presentations of Black versus White faces).

MEASUREMENT INSTRUMENTS

Measurement instruments in implicit social cognition are based on the idea that automatic responses are influenced by whatever mental contents are activated upon encountering a given object. Thus, when a mental association is sufficiently strong, activation of one concept can automatically spread to other associated concepts (Collins & Loftus, 1975), and thereby influence responses on the task. For example, if a person has strong associations with Coca-Cola, seeing a can of Coca-Cola should activate these associations automatically, which should influence the person's responses to stimuli that are conceptually congruent or incongruent with these associations. Most measurement instruments in implicit social cognition make use of this logic in one way or another (Moors, Spruyt, & De Houwer, 2010).

Sequential Priming Tasks

The first type of measurement instruments in the area of implicit social cognition is based on the logic of sequential priming. (For a review, see Wentura & Degner, 2010.) In a typical sequential priming task, participants are briefly presented with a prime stimulus, which is followed by a target stimulus. Depending on the nature of the task, participants are asked to (1) classify the target as positive or negative (i.e., evaluative decision task), (2) classify the target in terms of a semantic property (i.e., semantic decision task), or (3) decide whether the target is a meaningful word or a meaningless letter string (i.e., lexical decision task). The basic idea underlying sequential priming tasks is that quick and accurate responses to the target should be facilitated when the target is conceptually congruent with the associations that were activated by the prime stimulus. In contrast, quick and accurate responses to the target should be impaired when the target is conceptually incongruent with the associations that were activated by the prime stimulus.

For example, if a person has strong positive associations with Pepsi, this person should be faster and more accurate in identifying the valence of positive words when he or she has been primed with the word "Pepsi" compared to priming trials with a neutral baseline stimulus (Fazio, Sanbonmatsu, Powell, & Kardes, 1986). Conversely, evaluative classifications of negative words should be slower and less accurate when the person has been primed with the word "Pepsi" compared to priming trials with a neutral baseline stimulus. Different from the focus on evaluative associations in sequential paradigms with evaluative decision tasks, sequential priming with semantic decision tasks is used to measure semantic associations.

For example, a person with strong gender-stereotypic associations should show better performance in identifying the gender of female pronouns after being presented with stereotypically female professions (e.g., nurse) than stereotypically male professions (e.g., doctor), and vice versa (Banaji & Hardin, 1996). Last, using a lexical decision task to assess racial stereotypes, a person may show facilitated classifications of target words related to positive and negative stereotypes of African Americans (e.g., athletic, criminal) after being primed with Black faces compared to priming trials with a neutral baseline stimulus (Wittenbrink, Judd, & Park, 1997).

Sequential priming tasks have been used with supraliminally (e.g., Fazio et al., 1995) as well as subliminally presented primes (e.g., Wittenbrink et al., 1997). However, although widely used, sequential priming tasks have been criticized for their low reliability, which rarely exceed Cronbach's alpha values of .50 (Gawronski & De Houwer, 2014). This limitation has led researchers to develop alternative instruments that show reliability estimates that are comparable to the ones of traditional self-report measures.

Implicit Association Test

The most prominent example of such measures is the Implicit Association Test (IAT), which has been developed to overcome the known limitations of sequential priming tasks (Greenwald, McGhee, & Schwartz, 1998). In the critical blocks of the IAT, participants are asked to complete two binary categorization tasks that are combined in a manner that is either congruent or incongruent with the content of the to-be-measured attribute. For example, in the commonly used race IAT, participants may be asked to categorize pictures of Black and White faces in terms of their race and positive and negative words in terms of their valence. In one critical block of the task, participants are asked to press one response key for Black faces and negative words and another response key for White faces and positive words (i.e., prejudice-congruent block). In the other critical block, participants are asked to complete the same categorization tasks with a reversed key assignment for the faces, such that they have to press one response key for White faces and negative words and the other response key for Black faces and positive words (i.e., prejudice-incongruent block). The basic idea underlying the IAT is that responses in the task should be facilitated when two mentally associated concepts are mapped onto the same response key. For example, a person who has more favorable associations with Whites than Blacks should show faster and more accurate responses when White faces share the same response key with positive words and Black faces share the same response key with negative words, compared with the reversed mapping.

IAT scores are inherently relative in the sense that they conflate four conceptually independent constructs. For example, in the race IAT, a participant's performance is jointly determined by the strength of White-positive, Black-positive, White-negative, and Black-negative associations. This limitation makes the IAT inferior to sequential priming tasks, which permit the calculation of separate priming scores if the tasks include appropriate baseline primes (see Wentura & Degner, 2010). Yet the IAT is superior in terms of its internal consistency, which is typically in the range of .70 to .90 (Gawronski & De Houwer, 2014). The latter characteristic has contributed to it being the most widely used measurement instrument in implicit social cognition.

At the same time, the IAT has been criticized for its blocked presentation of "congruent" and "incongruent" trials, which has

been linked to several sources of systematic measurement error. For example, previously trained key mappings have been shown to influence performance in the second pairing in an IAT, such that IAT scores may differ depending on whether prejudice-congruent or prejudice-incongruent blocks are completed first (see Teige-Mocigemba, Klauer, & Sherman, 2010). To address these and various other limitations, researchers have developed several variants of the standard IAT that avoid blocked presentations of congruent and incongruent trials, permit nonrelative measurements for individual targets and attributes, and reduce the overall length of the task. Examples of these IAT variants include the Recoding-Free IAT (IAT-RF; Rothermund, Teige-Mocigemba, Gast, & Wentura, 2009), the Single-Block IAT (SB-IAT; Teige-Mocigemba, Klauer, & Rothermund, 2008), the Single-Category IAT (SC-IAT; Karpinski & Steinman, 2006), the Single-Attribute IAT (SA-IAT; Penke, Eichstaedt, & Asendorpf, 2006), and the Brief IAT (BIAT; Sriram & Greenwald, 2009).

Affect Misattribution Procedure

The affect misattribution procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005) was designed to combine the structural advantages of sequential priming tasks with the superior psychometric properties of the IAT. (For a review, see Payne & Lundberg, 2014.) Two central differences to traditional priming tasks are that (1) the target stimuli in the AMP are evaluatively ambiguous, and (2) participants are asked to report their subjective evaluations of the targets. That is, rather than inferring evaluative associations from the response time it takes a participant to decide whether a target stimulus is positive or negative, participants are presented with a neutral target stimulus and are asked to evaluate it. The basic idea is that participants may misattribute the affective feelings elicited by primes to the neutral targets and therefore judge the targets more favorably when they were primed with a positive stimulus than when they were primed with a negative stimulus. For example, in an AMP to measure racial attitudes, participants may be asked to indicate whether they find Chinese ideographs visually more pleasant or visually less pleasant than average after being primed with pictures of Black versus White faces. A preference for Whites over Blacks would be indicated by a tendency to evaluate the Chinese ideographs more favorably when the ideographs followed the presentation of a White face than when they followed the presentation of a Black face. Interestingly, priming effects in the AMP emerge even when participants are explicitly informed about the nature of the task and instructed not to let the prime stimuli influence their evaluations of the targets (Payne et al., 2005).

The AMP has been criticized for being susceptible to intentional use of the primes in evaluations of the targets (Bar-Anan & Nosek, 2012). However, the basis of this criticism has been refuted by research showing that relations between AMP effects and self-reported intentions to use the primes are due to retrospective confabulations of intentionality (i.e., participants infer that they must have had such intentions when asked afterward) rather than actual effects of intentional processes (e.g., Gawronski & Ye, 2015; Payne et al., 2013). The AMP was originally designed to measure evaluative associations, but newer versions have been developed to capture semantic associations (e.g., Sava et al., 2012).

Other Instruments

The procedures just described are the most commonly used instruments in implicit social

cognition. Yet several other instruments have been developed to address specific limitations of existing tasks. We briefly describe these procedures here. (For a comprehensive review and discussions of advantages and disadvantages of different procedures, see Gawronski & De Houwer, 2014.) Many of these procedures were designed to overcome specific limitations of the IAT (e.g., relative scores, blocked structure) while preserving its psychometric advantages.

In the go/no-go association task (GNAT; Nosek & Banaji, 2001) participants are presented with different kinds of stimuli sequentially and asked to press a button ("go") in response to two types of stimuli (e.g., positive words and White faces), and to withhold a reaction ("no go") in response to all other stimuli (e.g., negative faces and non-White faces). Participants typically are given a very short response window (e.g., 600 ms), and GNAT scores are calculated in terms of accuracy (rather than response times) using signal detection theory (Green & Swets, 1966). A major advantage of the GNAT is the possibility to calculate nonrelative scores for individual target objects (e.g., attitudes toward Blacks) instead of relative scores involving two target objects (e.g., relative preference for Whites of Blacks). However, the GNAT has shown lower reliability estimates compared with the standard IAT (Gawronski & De Houwer, 2014).

On the Extrinsic Affective Simon Task (EAST; De Houwer, 2003) participants are presented with target words that are shown in two different colors (e.g., yellow and. blue) and positive and negative words in white color. In the critical block of the task, participants are asked to respond to positive white words and words of one color (e.g., yellow) with the same key and to negative white words and words of the other color (e.g., blue) with another key (or vice versa). Because the target words are presented in different colors over the course of the task, each target is sometimes paired with the response key for positive words and sometimes with the response key for negative words. The critical question is whether participants respond faster and more accurately to the targets depending on whether they require a response with the "positive" or the "negative" key. Although the EAST eliminates the block structure of the IAT and permits a calculation of nonrelative scores for individual target objects, it has been shown to be inferior to the IAT in terms of its reliability and construct validity. This limitation has been attributed to the feature that participants do not have to process the semantic meaning of the target words (De Houwer & De Bruycker, 2007b). To address this limitation, De Houwer and De Bruycker (2007a) have developed a modified variant of the EAST that ensures semantic processing of the target words, which they called the Identification-EAST (ID-EAST).

Approach-avoidance tasks make use of the idea that positive stimuli elicit approach reactions, whereas negative stimuli elicit avoidance reactions (e.g., Brendl, Markman, & Messner, 2005; Krieglmeyer & Deutsch, 2010; Schnabel, Banse, & Asendorpf, 2006). For example, Chen and Bargh (1999) found that participants were faster at pushing a lever toward themselves (approach) in response to positive as opposed to negative stimuli. Conversely, participants were faster at pushing a leaver away from themselves (avoidance) for negative as opposed to positive stimuli (cf. Solarz, 1960). In the area of implicit social cognition, such congruency effects have been utilized to assess spontaneous responses toward a variety of objects, including social groups (e.g., Neumann, Hülsenbeck, & Seibt, 2004) and food stimuli (e.g., Seibt, Häfner, & Deutsch, 2007). In contrast to early accounts that interpreted these effects in terms of direct links between

particular motor actions and motivational orientations (e.g., contraction of arm extensor = avoidance; contraction of arm flexor muscle = approach), recent research suggests that congruency effects in approach-avoidance tasks depend on the evaluative meaning that is ascribed to a particular motor action (e.g., Eder & Rothermund, 2008). Hence, responses toward the same stimuli (e.g., pulling a lever) can be reversed when the same movements are coined in negative terms (e.g., "downward") as opposed to positive terms ("pull"), and vice versa (e.g., "upward" versus "push").

In the sorting paired features task (SPFT; Bar-Anan, Nosek, & Vianello, 2009), participants are presented with pairs of stimuli (instead of just one) and provided with four instead of two response options that represent all possible combinations of stimulus types (e.g., White–good, White–bad, Black–good, and Black–bad). These response options are presented in the four corners of a computer screen and mapped onto four buttons on a computer keyboard. The specific location of the four response options is randomized over four blocks of the task. Participants' task is to quickly press the response key that captures the displayed pair of stimuli (e.g., press the key for Black-good in response to a Black face paired with a positive word). Scores are conceptualized as the difference in the response latency of accurately identifying a given combination compared to the other three combinations, standardized by each participant's individual response times across all trials. This algorithm allows for the calculation of individual rather than relative scores.

The Action Interference Paradigm (AIP; Banse, Gawronski, Rebetez, Gutt, & Morton, 2010) has been developed for research with young children for whom the demands of existing tasks might be too overwhelming. For example, using a variant of the AIP to measure gender stereotypes, Banse et al. (2010) asked children to distribute gender-stereotypical gifts (i.e., trucks and dolls) to boys and girls by pressing one of two buttons that were marked with images of a boy and a girl. In one block of the task, the children were told that the boy would like to get a truck and the girl would like to get a doll (i.e., stereotype-congruent block). In another block of the task, the children were told that the boy would like to get a doll and the girl would like to get a truck (i.e., stereotype-incongruent block). The AIP uses response latencies to measure the ease of responding similar to the IAT. Although the AIP has been developed specifically to measure gender stereotypes, procedural modifications could make it amenable for the assessment of other constructs (Gawronski & De Houwer, 2014).

Deviating from the concern with measuring associations between concepts, the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010) has been developed to measure propositional representations that capture how concepts are related. For example, in an IAT to measure self-esteem (Greenwald & Farnham, 2000), facilitated responses in the block that combines self-related words and positive words may reflect a person's actual self (i.e., *I am good*), but it may also reflect the person's ideal self (i.e., *I want to be good*). Research by Remue, Hughes, De Houwer, and De Raedt (2014) has shown that the two kinds of underlying representations are indeed conflated in the standard IAT, which can lead to theoretically implausible results (e.g., high levels of implicit self-esteem among depressed participants; see De Raedt, Schacht, Franck, & De Houwer, 2006). To overcome this limitation, the procedure includes presentations of two stimuli, such as a target object (e.g., *me*) and a valenced word (e.g., *good*). The response

keys are labeled to describe different ways in which the two stimuli are related (e.g., similar versus different). Across several blocks of the task, participants are trained to learn that one key is the correct one for one type of combination and the other key is the correct one for the opposite combination. For example, participants might be trained to press the "similar" key when they are presented with the stimulus pair *I am* and *good* and the "different" key when they are presented with the stimulus pair *I am* and *bad*, or vice versa. Alternatively, participants might be trained to press the "similar" key when they are presented with the stimulus pair *I want to be* and *good* and the "different" key when they are presented with the stimulus pair *I want to be* and *bad*, or vice versa. The basic idea underlying the Implicit Relational Assessment Procedure is that responses in the task should be facilitated when a person's representation is congruent with the relation captured by the required response key than when it is incongruent with the required response. (For an alternative measure capturing propositional representations, see De Houwer, Heider, Spruyt, Roets, & Hughes, 2015).

RELATIONSHIP BETWEEN IMPLICIT AND EXPLICIT MEASURES

A common rationale for the use of implicit measures is that they provide information that cannot be captured by explicit measures. This argument is based on the observation that implicit and explicit measures tend to be weakly related. Hofmann, Gawronski, Gschwendner, Le, and Schmitt (2005) conducted a meta-analysis on the relation between IAT scores with corresponding self-reports and found an average correlation of .24. Cameron, Brown-Iannuzzi, and Payne (2012) found similar results in a meta-analysis on sequential priming tasks. However, in both cases there was also considerable variation in correlations, depending on the domain studied as well as procedural and methodological factors. Overall, correlations between implicit and explicit measures tend to be larger for self-reported judgments of feelings and affect compared to more cognitive judgments (e.g., Gawronski & LeBel, 2008; Smith & Nosek, 2011). For example, in a study by Banse, Seise, and Zerbes (2001), scores of a gay–straight IAT showed higher correlations to self-reported affective reactions toward gay people (e.g., self-reported affect when seeing two men kissing each other) compared to self-reported cognitive reactions (e.g., agreement with the statement that gay men should not be allowed to work with children). Implicit and explicit measures also show higher correlations when participants are given less time to think about their judgments than when they are encouraged to deliberate about their response (e.g., Ranganath, Smith, & Nosek, 2008). Concerning method-related factors, correlations are generally higher when implicit and explicit measures are matched in terms of their dimensionality and content. For example, implicit measures reflecting relative preferences for one group over another tend to show higher correlations to explicit measures of the same relative preference compared to explicit measures of absolute evaluations (e.g., Hofmann et al., 2005). Similarly, implicit measures reflecting evaluations of Black and White faces typically show higher correlations to explicit measures using the same faces compared to explicit evaluations of antidiscrimination policies (e.g., Payne, Burkley, & Stokes, 2008).

Different theories have been proposed to explain variations in the relation between implicit and explicit measures, two of which will be described here. Although both

theories were formulated to explain relations between implicit and explicit measures in the area of attitudes, their basic assumptions can be applied to nonevaluative domains as well. (For a review, see Hofmann, Gschwendner, Nosek, & Schmitt, 2005.)

The MODE model (*M*otivation and *O*pportunity as *DE*terminants) assumes that implicit measures capture the automatic activation of attitudes in response to an object (Fazio, 2007). Depending on a person's motivation and opportunity, the person may engage in deliberate processing to scrutinize specific attributes of the object. In this case, people are assumed to base their judgments on the nature of relevant attributes instead of the automatically activated attitude. Hence, to the extent that both the motivation and the opportunity for deliberate processing are high, correlations between implicit and explicit evaluations are predicted to be low. Yet, when either the motivation or the opportunity for deliberate processing are low, people are assumed to rely on their automatic reactions, leading to higher correlations between implicit and explicit measures. These assumptions are supported by several studies indicating that evaluative judgments provided under time pressure show higher correlations with implicit measures compared to judgments provided without time pressure (e.g., Ranganath et al., 2008). Further evidence for the MODE model comes from research showing that participants with high motivation to control prejudice show lower correlations between implicit and explicit measures of racial prejudice compared to participants with a low motivation to control prejudice (e.g., Fazio et al., 1995).

Another theory that explains the relation between implicit and explicit measures is the associative-propositional evaluation (APE) model (Gawronski & Bodenhausen, 2006, 2011). According to the APE model, implicit measures capture the behavioral outcomes of associative processes; explicit measures are assumed to reflect the behavioral outcomes of propositional processes. Associative processes are defined as the activation of mental associations on the basis of feature similarity and spatiotemporal contiguity; propositional processes are defined as the validation of the information implied by activated associations. A central assumption of the APE model is that the propositional validation of activated associations involves an assessment of consistency, in that inconsistency requires a reassessment and potential revision of one's beliefs (Gawronski, 2012). Thus, correspondence between implicit and explicit measures is assumed to depend on whether the association captured by an implicit measure is consistent with other information that is considered for a self-reported judgment. To the extent that it is consistent with other salient information, it is usually regarded as valid and therefore used as a basis for self-reported judgments. However, if it is inconsistent with other salient information, people may reject this association in order to restore cognitive consistency (Gawronski & Strack, 2004).

Although the MODE and the APE model make similar predictions in most cases, the theories differ in terms of two central assumptions. First, whereas the MODE model assumes that motivation and opportunity are the primary determinants of implicit–explicit relations, the APE model proposes cognitive consistency as the central proximal factor. To illustrate this difference, consider Fazio et al.'s (1995) finding that the relation between implicit and explicit measures of prejudice is higher for participants with low motivation to control prejudice compared to participants with high motivation to control prejudice. From the perspective of the APE model, implicit measures of prejudice capture the affective reaction that results

from the associations that are activated in response to members of the target group (e.g., negative affective reaction to African Americans resulting from negative associations). This reaction may serve as the basis for a self-reported evaluative judgment (e.g., *I dislike African Americans*), unless such a judgment would be inconsistent with other salient information. In the case of racial prejudice, other salient information may include egalitarian beliefs (e.g., *Negative evaluations of disadvantaged groups are wrong*) and beliefs about discrimination (e.g., *African Americans represent a disadvantaged group*). According to APE model, consistency among these beliefs may be restored by rejecting one's affective reaction as a basis for a self-reported evaluative judgment (e.g., *I like African Americans*). Yet consistency may also be restored by changing one's egalitarian beliefs (e.g., *Negative evaluations of disadvantaged groups are okay*) or one's beliefs about discrimination (e.g., *African Americans do not represent a disadvantaged group*). These considerations lead to the novel prediction that strong egalitarian beliefs (i.e., high motivation to control prejudice) should be insufficient to reduce the relation between implicit and explicit measures of racial prejudice when participants maintain cognitive consistency by denying racial discrimination. In this case, a person may report negative feelings toward African Americans and nevertheless maintain the belief that one should not express negativity toward disadvantaged groups, because the person denies that African Americans represent a disadvantaged group (akin to the concept of "modern racism"; McConahay, 1983). This prediction has been confirmed by Gawronski, Peters, Brochu, and Strack (2008), who found high correlations between implicit and explicit measures of racial prejudice when either egalitarian beliefs or perceived discrimination were low. Correlations between the two measures were reduced only when both egalitarian beliefs and perceived discrimination were high (see also Brochu, Gawronski, & Esses, 2011). These results suggest that cognitive consistency functions as the primary proximal determinant of implicit–explicit relations, whereas motivation and opportunity are better understood as distal determinants.

Second, whereas the MODE model assumes that deliberate processing generally reduces the relation between implicit and explicit measures, the APE model assumes that such reductions should occur only when the additionally considered information is inconsistent with the association captured by the implicit measure. To the extent that deliberate processing involves a selective search for information that supports the validity of this association, deliberate processing may in fact increase rather than decrease the relation between implicit and explicit measures. This hypothesis is consistent with research showing that selective search for information that is consistent with activated associations increases the correlation between implicit and explicit measures (e.g., Galdi, Gawronski, Arcuri, & Friese, 2012; see also Peters & Gawronski, 2011b).

PREDICTION OF BEHAVIOR

A major line of research in implicit social cognition aims to improve our understanding of psychological phenomena by using implicit measures to predict meaningful psychological outcomes (e.g., interpersonal behavior, decisions, mental health). Although the practical implications of the observed effect sizes has been the subject of debate (e.g., Greenwald, Banaji, & Nosek, 2015; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013), recent meta-analyses tend to support the predictive

validity of implicit measures (e.g., Cameron et al., 2012; Greenwald, Poehlman, Uhlmann, & Banaji, 2009). According to Perugini, Richetin, and Zogmeister (2010), implicit measures may contribute to the prediction of psychological outcomes over and above explicit measures in various ways, including (1) additive patterns, (2) double-dissociation patterns, (3) moderation patterns, and (4) interactive patterns.

Additive patterns involve cases in which implicit and explicit measures of the same construct jointly predict a particular outcome. Such cases tend to emerge when implicit measures are able to capture particular aspects of the outcome that are not captured by the explicit measure. For example, in a study on the prediction of consumer behavior, Maison, Greenwald, and Bruin (2004) found that adding an implicit measure of brand preferences increased the prediction of consumer choices over and above explicit measures.

Although additive patterns have been obtained in a few studies, a more common finding is a double dissociation in the prediction of different kinds of outcomes. Many dual-process models conceptualize implicit and explicit measures in terms of different underlying processes (e.g., Fazio, 2007; Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006; Strack & Deutsch, 2004). Based on this idea, implicit measures have been claimed to be superior in the prediction of spontaneous behavior, whereas explicit measures are assumed to be superior in the prediction of deliberate behavior. (For a review, see Friese, Hofmann, & Schmitt, 2008.) In line with these assumptions, nonverbal behavior in interracial interactions has shown stronger relations with implicit as compared to explicit measures, whereas verbal behavior has been shown to reveal stronger relations to explicit as compared to implicit measures (e.g., Dovidio, Kawakami, & Gaertner, 2002). Similar findings have been obtained for the self-concept of shyness (Asendorpf, Banse, & Mücke, 2002), showing that implicit measures outperformed explicit measures in the prediction of spontaneous behavior (e.g., body posture), whereas explicit measures outperformed implicit measures in the prediction of deliberate behaviors (e.g., speech duration).

Despite the available evidence for double-dissociation patterns, several studies have shown only partial or weak dissociation patterns (Perugini et al., 2010). In these cases, implicit measures predicted spontaneous behavior and explicit measures predicted deliberate behavior, but either or both measures also predicted the respective other behavior (e.g., Richetin, Perugini, Adjali, & Hurling, 2007). From the perspective of dual-process theories, these patterns might be due to the fact that many behaviors are not cleanly classifiable as either spontaneous or deliberate but might instead have both spontaneous and deliberate elements. Thus, partial-dissociation patterns might be better described as a mixture of both additive and double-dissociation patterns in the prediction of outcomes. Based on these considerations, Perugini et al. (2010) suggested that these patterns may also be called partial additive patterns (when one measure predicts both kinds of behaviors, but the other measure predicts only one) or double additive patterns (when both implicit and explicit measures predict both spontaneous and deliberate behaviors).

Drawing on the assumptions of dual-process theories (e.g., e.g., Fazio, 2007; Strack & Deutsch, 2004), several studies have investigated factors that determine whether the same outcome is predicted by either implicit or explicit measures. Such findings can be described as reflecting a moderation pattern. The central idea underlying this research is that aspects of the person

or the situation can influence the degree of control over a given behavior, which should determine whether the behavior is predicted better by either explicit or implicit measures. (For a review, see Friese et al., 2008.) Consistent with this idea, implicit measures have been shown to outperform explicit measures in the prediction of candy consumption when participants' cognitive resources were depleted. In contrast, explicit measures outperformed implicit measures in the prediction of candy consumption under control conditions, where participants presumably devoted their cognitive resources to controlling their eating behavior (e.g., Hofmann, Rauch, & Gawronski, 2007). Parallel findings have been obtained for individual differences in working memory capacity (WMC), such that eating behavior was predicted better by implicit measures for participants with low WMC, whereas the same behavior was better predicted by explicit measures for participants with high WMC (e.g., Hofmann, Gschwendner, Wiers, Friese & Schmitt, 2008). Together, these findings demonstrate how both individual differences and situational factors can determine whether implicit or explicit measures are superior in the prediction of a given behavior.

Deviating from approaches in which implicit and explicit measures are seen as competitors in the prediction of behavior, several studies have investigated interactive relations between the two kinds of measures. The central assumption underlying these studies is that discrepancies between implicit and explicit measures are indicative of an unpleasant psychological state that people aim to reduce (Rydell & McConnell, 2010). In line with this assumption, Rydell, McConnell, and Mackie (2008) found that participants who had been experimentally induced to hold discrepant evaluations of a fictitious target person on implicit and explicit measures scrutinized persuasive arguments from this person more thoroughly than participants who were induced to hold convergent evaluations. In general, people who show large discrepancies on implicit and explicit measures of a particular psychological attribute (e.g., attitude, self-concept) have been shown to process discrepancy-related information more extensively than people with small discrepancies (see also Briñol, Petty, & Wheeler, 2006). Similarly, combinations of high self-esteem on explicit measures and low self-esteem on implicit measures have been shown to predict various kinds of defensive behaviors (e.g., Jordan, Spencer, Zanna, Hoshino-Browne, & Correll, 2003). The basic idea behind this work is that such self-esteem discrepancies reflect a conflict between spontaneous feelings and deliberate thoughts about the self, which leads to a threatening state of insecurity that people try to overcome through various kinds of defensive behaviors (e.g., narcissistic tendencies, increased in-group bias).

Despite the available evidence for each of the four patterns, their boundary conditions are still not well understood (Perugini et al., 2010). Although many of the original predictions regarding the different patterns were derived from dual-process theories, specific predictions regarding the conditions under which each of them should occur are still lacking. Thus, an important task for future research is to identify the boundary conditions of different predictive patterns and to develop theories that explain why their occurrence depends on the identified conditions.

FORMATION, CHANGE, AND CONTEXT EFFECTS

Another central question in implicit social cognition concerns the situational determinants of variations on implicit measures.

We divide our discussion of this work into three parts that address distinct theoretical questions: (1) factors that influence the formation of mental representations, (2) factors that lead to changes in existing mental representations, and (3) context effects on the activation of existing representations. We also discuss (4) the lack of process purity of implicit measures, suggesting that some variations may be due to factors that are unrelated to the constructs of interest.

Formation

Theoretically, variations on implicit measures are best understood as reflecting the formation of a new mental representation when (1) the target object is unknown to participants and (2) the acquisition of novel information about the target object causes systematic variations on implicit measures. Empirical evidence suggests that such variations can be caused by descriptive information about an object (often called propositional learning) as well as repeated pairings between a target object and other stimuli (often called associative learning; for a review, see Gawronski & Sritharan, 2010).

The simplest example for effects of descriptive information comes from studies in which participants were given positive or negative information about unknown objects, individuals, or groups (e.g., Gregg, Seibt, & Banaji, 2006). Such effects have been shown for as a little as three statements (e.g., Gawronski, Walther, & Blank, 2005). Theoretically, these findings contradict the widespread assumption that implicit measures reflect highly overlearned associations that result from long-term socialization experiences (e.g., Rudman, 2004; Wilson, Lindsey, & Schooler, 2000). Although there is evidence that developmental factors can contribute to variations on implicit measures (e.g., Baron & Banaji, 2006; Rudman, Phelan, & Heppen, 2007), research demonstrating such rapid effects on implicit measures prohibit the reverse conclusion that variations on implicit measures could be interpreted as indicators of early life experiences.

The idea of associative learning is most prominently reflected in research on evaluative conditioning (EC). In EC, repeated pairings of a neutral conditioned stimulus (CS) with a positive or negative unconditioned stimulus (US) lead to changes in the evaluation of the CS in line with the valence of the US. (For a review, see De Houwer, Thomas, & Baeyens, 2001; for a meta-analysis, see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010.) EC effects have also been demonstrated on implicit measures (e.g., Olson & Fazio, 2001). The central assumption underlying this research is that CS–US pairings create new associations in memory, which can be captured by implicit measures.

Associative processes have also been implicated in other effects that do not involve the presentation of repeated pairings. For example, investigating effects of mere ownership with implicit measures, Gawronski, Bodenhausen, and Becker (2007) found that participants showed more favorable evaluations of newly owned objects compared to objects that they did not own. According to Gawronski et al., such ownership effects are due to a process called associative self-anchoring. The central feature of this process is that a newly owned object becomes mentally associated with the self, which leads to an associative transfer of one's self-evaluation to the owned object. To the extent that most people hold positive evaluations of themselves (e.g., Bosson, Swann, & Pennebaker, 2000; Greenwald & Farnham, 2000; Koole, Dijskterhuis, & van Knippenberg, 2001), newly owned objects should elicit more favorable responses, and

these responses can be captured by implicit measures. Following a similar logic, the finding that in-groups are evaluated more favorably on implicit measures compared to out-groups has been attributed to the formation of associations between the self and one's in-group, which should lead to an associative transfer on self-evaluations to one's in-group (e.g., Roth & Steffens, 2014).

Change

Theoretically, variations on implicit measures can be understood as reflecting changes in existing mental representations when (1) the target object is well-known to participants and (2) the acquisition of new information about the target object causes systematic variations on implicit measures. In the early years of implicit social cognition, implicit measures were claimed to be much more resistant to change than explicit measures (e.g., Rudman, 2004; Wilson et al., 2000). However, this assumption has been refuted by numerous studies showing changes on implicit measures in the absence of changes on explicit measures (e.g., Gawronski & LeBel, 2008; Olson & Fazio, 2006). At the same time, there have been several demonstrations of changes on explicit measures in the absence of changes on implicit measures (e.g., Gawronski & Strack, 2004; Gregg et al., 2006). Thus, a central question in this line of research is when the acquisition of new information leads to (1) changes on implicit but not explicit measures, (2) changes on explicit but not implicit measures, and (3) corresponding changes on both explicit and implicit measures.

According to the APE model (Gawronski & Bodenhausen, 2006, 2011), changes on implicit but not explicit measures should occur when (1) a given factor influences the structure and content of associations in memory and, at the same time, (2) these newly created associations are rejected as a basis for self-reported judgments because of their inconsistency with other salient information. Resonating with the idea of associative learning in EC, this pattern has been observed most commonly when (1) a well-known CS has been repeatedly paired with a positive or negative US, presumably leading to the formation of new associations, but (2) participants rely on other information that leads them to reject the newly formed associations as a basis for their evaluative judgments of the CS (e.g., Gawronski & LeBel, 2008; Gibson, 2008; Grumm, Nestler, & von Collani, 2009; Karpinski & Hilton, 2001; Olson & Fazio, 2006). However, when participants were encouraged to rely on their affective feelings toward the CS, implicit and explicit measures typically showed corresponding effects, in that both reflected the valence of the CS–US pairings (e.g., Gawronski & LeBel, 2008; Grumm et al., 2009). The latter finding is consistent with the APE model's prediction that both implicit and explicit measures should show change when (1) a given factor influences the structure and content of associations in memory and (2) these newly created associations are accepted as a valid basis for self-reported judgments.

Another prediction of the APE model (Gawronski & Bodenhausen, 2006, 2011) is that changes on explicit but not implicit measures should occur when (1) a given factor influences the perceived validity of associations in memory and, at the same time, (2) this factor does not result in the formation of new associations. According to the APE model, this case is most likely when newly acquired information leads to inconsistency within a set of salient beliefs, and the resulting inconsistency is resolved by rejecting activated associations as a basis for self-reported judgments. Consistent with these assumptions, research by Gawronski

and Strack (2004) has shown that cognitive dissonance arising from induced compliance (cf. Festinger & Carlsmith, 1959) leads to changes on explicit but not implicit measures (see also Wilson et al., 2000). The same pattern has been observed in paradigms where previously acquired information is discredited as invalid, and participants are asked to mentally reverse the previously presented information. For example, Gregg et al. (2006) presented participants with positive information about a Group A and negative information about another Group B. Next, participants were told to mentally reverse this information, such that the positive information was supposed to refer to Group B and the negative information was supposed to refer to Group A. Whereas explicit measures showed a full reversal, implicit measures reflected the content of the initial information.

A critical aspect in these studies is that the discrediting information involves a simple "negation" of activated associations, which may lead to a rejection of these associations for a judgement. Yet mere rejection of a given association for overt judgments does not necessarily lead to a deactivation of this association (see Deutsch, Gawronski, & Strack, 2006). In fact, repeated negations may often have ironic effects, in that they strengthen the associative link that is supposed to be undone. For example, rejecting the proposition "Old people are bad drivers" as false may have counterintentional effects at the associative level, in that it may strengthen the associative link between *old people* and *bad drivers*. Consistent with this hypothesis, research found that repeated negations of a stereotype enhanced (rather than reduced) the stereotypical responses on implicit measures. A successful reduction occurred only when participants repeatedly affirmed a counterstereotype (e.g., Gawronski, Deutsch, Mbirkou, Seibt, & Strack, 2008). Effective changes of this kind have also been obtained in studies showing that novel evaluative information that is highly diagnostic (e.g., Cone & Ferguson, 2015) or suggests a reinterpretation of earlier information (e.g., Mann & Ferguson, 2015) effectively reverses responses on both explicit and implicit measures. According to the APE model, this pattern can be observed when (1) a given factor leads to a change in perceived validity of activated information and (2) new associations are formed by the process of propositional validation.

To summarize the different patterns that can emerge as a result of interactions between associative and propositional processes, Gawronski and Bodenhausen (2006) provided a schematic overview that includes four cases:

Case 1: A direct effect on associative representations with the newly formed associations being accepted by a propositional validity assessment. This pattern is assumed to lead to corresponding changes on implicit and explicit measures, with the change on the explicit measure being fully mediated by the change on the implicit measure (e.g., Gawronski & LeBel, 2008; Whitfield & Jordan, 2009).

Case 2: A direct effect on associative representations with the newly formed associations being rejected by a propositional validity assessment. This pattern is assumed to lead to changes on implicit but not explicit measures (e.g., Gawronski & LeBel, 2008; Olson & Fazio, 2006).

Case 3: A direct effect on the process of propositional validity assessment that leads to a rejection of activated associations. This pattern is assumed to lead to changes on

explicit but not implicit measures (e.g., Gawronski & Strack, 2004; Gregg et al., 2006).

Case 4: Acquisition of new propositional information that leads to the formation of new associations. This pattern is assumed to lead to corresponding changes on implicit and explicit measures, with the change on the implicit measure being fully mediated by the change on the explicit measure (e.g., Gawronski & Walther, 2008; Whitfield & Jordan, 2009).

Additionally, when a given situation involves multiple factors with different effects, the four basic cases can also occur in various combinations. For example, opposite effects on implicit and explicit measures have been observed when repeated CS–US pairings imply an evaluation that is opposite to the one implied by newly acquired propositional information. In such cases, implicit measures have been shown to reflect the evaluation implied by the CS–US pairings, whereas explicit measures reflect the valence of the newly acquired propositional information (e.g., Moran & Bar-Anan, 2013; Rydell, McConnell, Mackie, & Strain, 2006).

Context Effects

Theoretically, variations on implicit measures can be understood as reflecting contextually induced shifts on implicit measures when (1) the target object is known to participants and (2) contextually induced variations occur in the absence of new information about the target. Consistent with the idea of contextually induced shifts, a growing body of research has shown that implicit measures are highly malleable and context dependent. (For reviews, see Blair, 2002; Gawronski & Sritharan, 2010.) For example, in research using implicit measures of racial prejudice, White participants showed more positive evaluations of Black targets when the targets were presented in the context of a church than when they were presented in the context of graffiti wall (Wittenbrink, Judd, & Park, 2001). Similarly, Roefs et al. (2006) found that evaluations of high-fat foods on implicit measures were more favorable when the foods were presented in a restaurant context than when they were presented in the context of a health clinic. Similar effects have been obtained for a wide range of contextual factors, including recently encountered members of a social group (e.g., Dasgupta & Greenwald, 2001), social roles (e.g., Richeson & Ambady, 2003), salient categories (e.g., Mitchell, Nosek, & Banaji, 2003), and mood states (e.g., Gemar, Segal, Sagrati, & Kennedy, 2001).

The context dependence of implicit measures has fueled theoretical debates as to whether implicit measures reflect stable representations in memory (e.g., Fazio, 2007) or online constructions on the basis of momentarily accessible information (e.g., Schwarz, 2007). According to representational accounts, spontaneous responses captured by implicit measures depend on how a target object is categorized. To the extent that contextual cues influence the categorization of a given object, these cues may influence which category representation is activated in response to the object, which in turn influences spontaneous responses on implicit measures. In contrast, constructivist accounts propose that spontaneous responses on implicit measures depend on momentarily accessible attributes rather than on abstract category representations. Thus, to the extent that contextual cues influence the relative accessibility of certain attributes, these cues should lead to variations in a person's responses to the same object, which should be captured by implicit measures.

Although representational and constructivist accounts attribute context effects to fundamentally different processes, either one of them can explain the available evidence for context effects on implicit measures. However, their explanations may be criticized as circular, in that they can explain any context effect in a post hoc fashion without providing testable predictions about their boundary conditions. To address this concern, Gawronski, Rydell, Vervliet, and De Houwer (2010) proposed an integrative theory that combines components of both representational and constructivist accounts. A central aim of the theory is to provide a priori predictions about the contextual conditions under which implicit measures reflect (1) initially acquired information, (2) subsequently acquired information that is inconsistent with the initial information, or (3) a mixture of both. There are two core assumptions of the theory: (1) attention to contextual cues during the encoding of evaluative information determines whether this information is stored in a context-free or contextualized representation, and (2) attention to contextual cues is typically low during the encoding of initial information, but enhanced by exposure to expectancy-violating information. Together, the two assumptions imply that initial experiences tend to be stored in context-free representations, whereas expectancy-violating information is usually stored in contextualized representations.

Applied to context effects on implicit measures, Gawronski et al.'s (2010) theory predicts that implicit measures should reflect the content of expectancy-violating information only in the context in which this expectancy-violating information was learned; whereas they should reflect the content of initially acquired information in any other context. This includes both the context in which this information was originally acquired and any novel context in which the target object has not been encountered before. These predictions have been confirmed in a series of studies by Gawronski et al., which also tested several predictions about how attentional processes can moderate the hypothesized patterns of context effects (see also Gawronski, Ye, Rydell, & De Houwer, 2014; Rydell & Gawronski, 2009). In addition to providing precise predictions about the conditions under which implicit measures should be context dependent or context independent, another contribution of the theory is that it provides novel, empirically confirmed predictions about contextual conditions under which implicit measures should change in response to novel information and under which conditions they should be resistant to change. (For a review, see Gawronski & Cesario, 2013.) The central prediction is that change should be more likely when the target object is subsequently encountered in the context in which the new information was acquired. Yet change is less likely to occur when the target object is subsequently encountered in a context that is different from the one in which the new information has been acquired. These predictions have been confirmed in several independent studies and corroborated in a recent meta-analysis (Gawronski, Hu, Rydell, Vervliet, & De Houwer, 2015).

Lack of Process Purity

In the introductory section to this chapter, we noted the fundamental role of mental associations as a core concept of implicit social cognition. In line with this idea, implicit measures are often assumed to provide direct proxies for mental associations. However, in a strict sense, implicit measures reflect behavioral outcomes, and these outcomes should not be equated with their mental underpinnings (De Houwer, Gawronski, & Barnes-Holmes, 2013). Although the impact

of mental associations on implicit measures is rarely disputed in the field of implicit social cognition (for an exception, see De Houwer 2014), a considerable body of research suggests that implicit measures do not provide process-pure reflections of mental associations (Teige-Mocigemba, Klauer, & Sherman, 2010). To disentangle the contributions of multiple qualitatively distinct processes to implicit measures, theorists have developed formal models that provide quantitative estimates of these processes. These models include applications of process dissociation (Payne & Bishara, 2009), multinomial modeling (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Meissner & Rothermund, 2013; Stahl & Degner, 2007), and diffusion modeling (Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007).

One of the most prominent examples is Conrey et al.'s (2005) quad model, which distinguishes among four qualitatively distinct processes underlying responses on implicit measures: (1) activation of an association, (2) detection of the correct response required by the task, (3) success at overcoming associative bias, and (4) guessing. Research using the quad model has provided more fine-grained insights into the mechanisms underlying previous findings obtained with implicit measures. Whereas some effects have been shown to be genuinely related to underlying associations, others stem from nonassociative processes, such as the ability to inhibit activated associations. (For a review, see Sherman et al., 2008). For example, whereas extended training to associate racial groups with positive or negative attributes has been shown to influence associative bias (Calanchini, Gonsalkorale, Sherman, & Klauer, 2013), alcohol-related increases in implicit measures of racial bias have been linked to impaired inhibitory control (Sherman et al., 2008). Similarly, higher scores on implicit measures of racial bias among older adults have been shown to be related to decreased ability to control associations rather than stronger negative associations compared to younger adults (Gonsalkorale, Sherman, & Klauer, 2009).

QUESTIONS AND CONTROVERSIES

Much of the popularity of implicit measures can be explained by the promise that they provide insights that cannot be gained with explicit measures (e.g., when people are either unwilling or unable to provide accurate self-reports). However, although some claims have received empirical support, others have been challenged by an accumulating body of evidence. In this section, we discuss some frequent assumptions and ongoing controversies about what implicit measures do and do not tell us.

Do Implicit Measures Uncover Unconscious Representations?

As discussed earlier in this chapter, a central component in the historical origin of implicit social cognition has been Greenwald and Banaji's (1995) definition in terms of "introspectively unidentified (or inaccurately identified) traces of past experience that mediate responses" (p. 5). Although the original definition referred to unconscious sources of mental representations, it has often been interpreted in the sense that the mental representations themselves are unconscious. The latter interpretation has become so common that many authors describe the constructs captured by implicit measures as unconscious attitudes, unconscious prejudice, unconscious stereotypes, unconscious self-esteem, and unconscious self-concepts (e.g., Bosson et al., 2000; Cunningham, Nezlek, &

Banaji, 2004; Rudman, Greenwald, Mellott, & Schwartz, 1999). These constructs are contrasted with the ones captured by explicit measures, which are often described as conscious attitudes, conscious prejudice, conscious stereotypes, conscious self-esteem, and conscious self-concepts. Empirically, the claim that implicit measures uncover unconscious representations whereas explicit measures reflect conscious representations is typically based on the low correlations between implicit and explicit measures frequently observed in this line of work (see Cameron et al., 2012; Hofmann et al., 2005).

Although it is correct that correlations between implicit and explicit measures may be low if the representations captured by implicit measures are unconscious, this valid inference does not justify the reverse conclusion that low correlations indicate an effect of unconscious representations on implicit measures (Gawronski, Hofmann, & Wilbur, 2006). After all, correlations between the two kinds of measures can be low for various reasons that have nothing to do with unconsciousness, including the motivation and opportunity to engage deliberate processing (Fazio, 2007) and cognitive inconsistency of activated mental contents (Gawronski & Bodenhausen, 2006). In fact, the unconsciousness hypothesis is at odds with the findings of studies in which participants were asked to predict their measurement scores on implicit measures. Using multiple IATs capturing attitudes toward different social groups, Hahn, Judd, Hirsh, and Blair (2014) found that participants were able to predict the patterns of their IAT scores with a high level of accuracy. Importantly, predicted and actual IAT scores were highly correlated within-subjects, although traditional explicit measures showed the same low correlations with IAT scores that are typically observed in this area. These findings pose a challenge to the claim that implicit measures provide a window into unconscious representations. Yet they are consistent with theories that explain dissociations between implicit and explicit measures in terms of other processes that involve a rejection of conscious representations (e.g., Fazio, 2007; Gawronski & Bodenhausen, 2006). Neither of these theories attributes the misalignment of implicit and explicit measures to lack of awareness.

An important aspect in this context is the distinction between awareness of one's own response (introspective awareness) and awareness of how one's own response compares to the responses of other people (social awareness). To obtain a high correlation between predicted and actual measurement scores in a typical between-subjects design, participants have to know not only their own response but also where their response falls in the distribution of responses revealed by the other participants (Hahn & Gawronski, 2014; Hahn et al., 2014). Thus, a more stringent way to test the conscious versus unconscious nature of the mental representations underlying implicit measures is to investigate correlations between predicted and actual measurement scores using within-subjects designs with multiple target objects. Whereas within-subjects correlations reflect the unique role of introspective awareness in predicting a person's measurement scores (e.g., how much do I like bananas compared to oranges, apples, mangoes, etc.?), the size of between-subjects correlations is additionally influenced by social awareness (e.g., how much do I like bananas compared to the other participants in the study?). Thus, in addition to the fact that low correlations between implicit and explicit measures can be attenuated by various factors related to the processing of target information (e.g., Fazio, 2007; Gawronski & Bodenhausen, 2006), the correlations obtained in traditional

between-subjects designs may underestimate the actual degree of introspective awareness when participants lack social awareness. Consistent with this concern, Hahn et al. (2014) found that participants were particularly good at predicting their measurement scores for a given target in relation to other targets, indicating high introspective awareness (within-subjects correlations in the range of .50–.60). Yet their accuracy was substantially lower for the prediction of measurement scores for a given target in relation to other participants, indicating lower social awareness (between-subjects correlations in the range of .30). In sum, evidence that participants are able to predict the patterns of their responses when asked contradicts the notion that implicit measures capture consciously inaccessible contents. Yet the prerequisites and consequences of such accurate predictions are still unclear at this point and require further investigation.

Is the Difference Between Implicit and Explicit Measures Just a Matter of Social Desirability?

A common idea underlying the use of implicit measures is that they are resistant to biasing effects of social desirability. Especially in the realm of prejudice, it is often assumed that people try to adjust their responses to social norms instead of honestly reporting their thoughts and feelings about social groups (e.g., Fazio et al., 1995). Empirically, this assumption translates into two hypotheses about implicit and explicit measures (Gawronski, LeBel, & Peters, 2007). First, the correspondence between implicit and explicit measures should be moderated by social desirability, such that the correlation between implicit and explicit measures should decrease as a function of increasing social desirability concerns. Second, it should be difficult to impossible for participants to strategically influence their scores on implicit measures.

Research examining the first hypothesis has produced mixed results. On one hand, Nosek (2005) found a significant relation between self-presentational concerns and the magnitude of correlations between implicit and explicit evaluations across 57 different attitude objects. On the other hand, a meta-analysis by Hofmann et al. (2005) did not find any relation between implicit–explicit correlations and the level of social desirability associated with a given content domain. Research that used individual difference measures of socially desirable responding (e.g., Crowne & Marlow, 1960) also failed to find the predicted relation with the magnitude of implicit–explicit correlations (e.g., Egloff & Schmuckle, 2003; Hofmann, Gschwendner, & Schmitt, 2005). More supportive evidence comes from research that has investigated self-presentational concerns for particular content domains rather than domain-independent concerns with socially desirable responding. For example, in the domain of prejudice, several studies found lower correlations between implicit and explicit measures for participants with a high motivation to control prejudice than for participants with a low motivation to control prejudice (e.g., Degner & Wentura, 2008; Fazio et al. 1995; Gawronski, Geschke, & Banse, 2003; Payne et al., 2005). Yet, as we outlined earlier in this chapter, even this pattern is limited to certain conditions, in that motivation to control prejudice has been shown to reduce implicit–explicit correlations only for participants who perceive high levels of discrimination, but not for participants who perceive low levels of discrimination (e.g., Brochu et al., 2011; Gawronski et al., 2008). The latter finding suggest that, although motivational factors do influence the correspondence between implicit and explicit measures,

their impact is more distal and mediated by cognitive consistency as a proximal factor (see Gawronski & Bodenhausen, 2006, 2011). Thus, although motivational factors contribute to dissociations between implicit and explicit measures, this conclusion does not permit the opposite conclusion that dissociations between implicit and explicit measures generally reflect a bias of dishonest or socially desirable responding on explicit measures.

Research examining the second hypothesis has also produced mixed results. On one hand, several studies show that instructions to bias or "fake" one's responses do not affect the scores of implicit measures (e.g., Asendorpf, Banse, & Mücke, 2002; Egloff & Schmukle, 2002; Steffens, 2004). On the other hand, research using the quad model has shown that variations in measurement scores are significantly related to differences in the success of overcoming associative biases (Conrey et al., 2005), which has been linked to a variety of individual differences and contextual factors. (For a review, see Sherman et al., 2008.) Overall, the available evidence to date suggests that, although implicit measures are less susceptible to strategic influences than explicit measures, implicit measures are not entirely immune to strategic control. Yet such influences seem to depend on a number of conditions, such as the use of particular response strategies (e.g., Teige-Mocigemba & Klauer, 2008), sufficient time (e.g., Degner, 2009), and prior experience with the task (e.g., Fiedler & Bluemke, 2005).

An important issue in this context is whether strategic influences involve either reactive control of one's responses on the task or proactive control of the mental contents that influence one's responses on the task (see Gawronski, LeBel, et al., 2007). Most research on "faking" effects has focused on reactive control of overt responses. The overall conclusion that can be drawn from this research is that reactive control is difficult but not impossible. Interestingly, research on proactive control has typically found a strong susceptibility of implicit measures to intentional influences. For example, in one of the first studies on this question, Blair, Ma, and Lenton (2001) found reduced scores on a gender-stereotyping IAT for participants who were asked to think vividly about counterstereotypical exemplars. Expanding on this finding, Peters and Gawronski (2011b) showed that recall of specific autobiographical memories can influence self-concept scores on an introversion–extraversion IAT, and this effect emerged regardless of whether participants were directly instructed to recall specific memories or the content of recalled memories was manipulated by making certain memories more desirable.

Together, the available evidence suggests that responses on implicit measures can be influenced through proactive control strategies involving the intentional activation of specific mental contents. Yet reactive control of one's responses on the task seems to be more difficult and contingent on various boundary conditions. Together, these findings show that effective control of implicit measures requires more elaborate strategies than control of explicit measures. However, they contradict the simplified notion that implicit measures are generally immune to strategic influences.

Do Implicit Measures Capture a Person's True Beliefs or Just Cultural Associations?

Another frequent question about implicit measures is whether they reflect a person's true beliefs or just culturally shared associations. The former interpretation resonates with the idea that implicit measures are

less susceptible to strategic control than explicit measures (e.g., Fazio et al., 1995). The latter interpretation is based on the idea that implicit measures might be influenced by incidental aspects of one's cultural environment that are not reflected in explicit measures (e.g., Arkes & Tetlock, 2004). To evaluate the validity of the competing views, we deem it important to distinguish between a philosophical and an empirical aspect of the debate.

The philosophical aspect concerns the question of which type of behavior should be regarded as reflecting a person's true self. On one hand, there is the view that a person's true self is revealed when intentional control over one's responses fails. On the other hand, there is the equally plausible view that a person's true self is reflected in what the person consciously intends to do or say. Whereas the first interpretation equates the true self with uncontrolled behavior, the second interpretation equates the true self with intentionally controlled behavior. To the extent that implicit measures capture responses under conditions of limited control and explicit measures capture intentionally controlled responses, the two philosophical views have conflicting implications about whether either implicit or explicit measures reflect a person's true self (Gawronski, Peters, & LeBel, 2008). However, because the preference for either of the two interpretations is a matter of philosophical worldviews rather than empirical observation, any claims about the true self depend on one's subjective preference for one or the other view. Thus, even though responses on implicit measures clearly fall into the category of behavior with limited control, any depiction of implicit measures as revealing the true self are contingent on the subjectively preferred conceptualization of the true self.

The empirical aspect of the debate concerns the questions of whether implicit measures are influenced by culturally shared associations, and, if so, whether behavior is more strongly influenced by a person's endorsed beliefs or culturally shared associations. Both questions can be answered on the basis of research reviewed in this chapter. As for the first question, research on EC suggests that implicit measures are highly sensitive to incidental pairings in the environment even when explicit measures do not show any effect of the pairings (e.g., Karpinski & Hilton, 2001; Olson & Fazio, 2006). Importantly, whether the resulting associations also influence explicit measures has been shown to depend on both the consideration of other information about the target object and the consistency of this information with the newly formed associations (e.g., Gawronski & LeBel, 2008; Grumm et al., 2009). From this perspective, the apparent conflict between the two views does not map onto two distinct types of mental associations (e.g., personal associations versus cultural associations). Instead, the debate becomes obsolete, because the endorsement of mental associations depends on the processes involved in their use for making a judgment. Moreover, the reviewed research on the prediction of behavior suggests that mental associations can influence behavior even when they are rejected as a basis for judgments and decisions. Yet, as we noted in the preceding sections, their behavioral impact is moderated by various factors related to the type of behavior (e.g., Asendorpf et al., 2002), the conditions under which the behavior is performed (e.g., Hofmann et al., 2007), and individual characteristics of the person who is performing the behavior (e.g., Richetin et al., 2007). From this perspective, the presumed boundary between two kinds of associations becomes rather blurry and difficult to defend at a conceptual level (see Gawronski, Peters, & LeBel, 2008).

Do Implicit Measures Reflect Associative or Propositional Processes?

In the introduction, we outlined that implicit social cognition as a field has been shaped by the idea that many key constructs of social psychology (e.g., attitudes, prejudice, stereotypes, self-esteem, self-concept) can be conceptualized as mental associations in memory (e.g., Greenwald et al., 2002). Expanding on this idea, an influential assumption of many dual-process theories is that implicit measures capture the behavioral outcomes of associative processes, whereas explicit measures capture the behavioral outcomes of propositional processes (e.g., Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004). Associative processes involve the activation of mental associations on the basis feature similarity and spatiotemporal contiguity; propositional processes involve the validation of the information implied by activated associations on the basis of cognitive consistency. A central difference between the two kinds of processes is that (1) associations can be activated regardless of whether they are regarded as valid or invalid, whereas propositional reasoning is inherently concerned with the perceived validity of activated information, and (2) mental propositions capture the particular relation between objects and events whereas associations reflect mere co-occurrence information (e.g., A causes versus prevents B; A likes or dislikes B).

Although dual-process interpretations of implicit and explicit measures are very common in implicit social cognition, they have been criticized by proponents of single-process theories who argue that both implicit and explicit measures are outcomes of a single propositional process (e.g., De Houwer, 2014; Kruglanski & Gigerenzer, 2011). The most elaborate single-process account has been put forward by De Houwer (2014), who argued that implicit measures reflect the automatic formation and activation of mental propositions about the relation between events. To support this argument, De Houwer cited several studies showing that implicit measures can be influenced by verbal instructions and inferences (e.g., De Houwer, 2006; Gast & De Houwer, 2012) and are sensitive to information about how stimuli are related (e.g., Gawronski et al., 2005; Zanon, De Houwer, & Gast, 2012). According to De Houwer (2014), dissociations between implicit and explicit measures occur because implicit measures involve constrained processing conditions during the retrieval of information, not because they tap into two distinct processes or representations. Whereas some information may be activated quickly without requiring a lot of cognitive effort, other information may require time and cognitive resources to be retrieved from memory. Thus, whereas the former type of information should have a strong effect on implicit measures, the latter type of information may influence only explicit, but not implicit, measures. Similar ideas have been advanced by researchers who emphasize the temporal dynamics of information activation and information integration in the course of generating an evaluative response (e.g., Cunningham, Zelazo, Packer, & Van Bavel, 2007; Wojnowicz, Ferguson, Dale, & Spivey, 2009).

In evaluating the two competing accounts, it is important to clarify the specific assumptions about which they disagree (see Gawronski, Brannon, & Bodenhausen, 2017). A central issue in this context is that effects of propositional processes on implicit measures have been addressed explicitly by dual-process theories that allow for mutual interactions between associative and propositional processes. Although it is true that some dual-process theories postulate a one-to-one mapping between processes and measures

(e.g., Rydell & McConnell, 2006), other dual-process theories assume that propositional inferences can function as a distal determinant of implicit measures to the extent that they change the structure or momentary activation of associations in memory (e.g., Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004). From this perspective, effects of verbal instructions and inferences on implicit measures are consistent with dual-process accounts as long as they are in line with their assumptions about the conditions of such top-down effects. In fact, dual-process theories imply two very specific predictions about the conditions under which top-down effects of propositional inferences on associative processes should occur, allowing for direct tests between single-process and dual-process accounts.

First, dual-process theories predict that information about the validity of observed stimulus contingencies should influence only explicit measures, whereas implicit measures should reflect stimulus contingencies regardless of their perceived validity. This prediction stands in contrast to the one implied by single-process propositional theories, which suggest that both explicit and implicit measures should reflect the perceived validity of stimulus contingencies. Second, dual-process theories predict that information about the relation between two co-occurring stimuli (e.g., A causes versus prevents B; A likes versus dislikes B) should influence only explicit evaluations, whereas implicit measures should reflect the mere co-occurrence of stimuli regardless of their relation. To date, research on the first pair of competing hypotheses confirmed the predictions of dual-process propositional theories (e.g., Peters & Gawronski, 2011a), whereas research on the second pair of competing hypotheses has found empirical support for the predictions of single-process theories (e.g., Moran & Bar-Anan, 2013). Yet the obtained effects seem to depend on a number of boundary conditions that are not addressed by either the two theories (e.g., Gawronski et al., 2005; Moran, Bar-Anan, & Nosek, 2015). Thus, despite the centrality of associative theorizing in the history of implicit social cognition, one of the most central questions to date concerns the nature of the processes and representations underlying implicit and explicit measures, and their implications for the debate between single-process versus dual-process theories. (For a review, see Sherman, Gawronski, & Trope, 2014.)

CONCLUSION

Research using implicit measures has provided valuable insights for many areas in psychology. Yet, as we noted throughout this chapter, there are still a number of unresolved questions that need to be addressed. Current models are well suited to explain different patterns in the prediction of behavior, but they lack specific predictions about the conditions under which a given pattern should occur. Similarly, discussions on whether the functional properties of implicit and explicit measures can be better explained by dual-process or single-process assumptions would benefit from research on the presumed roles of associative and propositional processes. We expect that both questions will play a central role in future research using implicit measures.

Another important, yet rarely acknowledged, issue is that different measurement instruments rely on different processes for the assessment of psychological attributes (Gawronski & De Houwer, 2014). Whereas some tasks are based on response interference mechanisms that involve a resolution of response conflicts (e.g., IAT), other tasks involve a disambiguation of ambiguous

stimulus features (e.g., AMP). If responses on these tasks are treated as behaviors rather than as proxies for underlying mental constructs (De Houwer et al., 2013), a stronger focus on underlying mechanisms suggests that the prediction of behavior with implicit measures might depend on the overlap between the processes underlying responses on the task and the processes underlying the to-be-predicted behavior. For example, whereas the IAT may be a better predictor of behavior involving a resolution of response conflicts (e.g., inhibition of an impulse to pull the trigger of a gun in response to a Black man holding an object that is identified as harmless), the AMP may be a better predictor of behavior involving a disambiguation of ambiguous stimuli (e.g., tendency to misidentify an ambiguous object as a gun when it is held by a Black man). Hence, in addition to shedding light on the contribution of multiple distinct processes to overt responses on the tasks, a stronger focus on the mechanisms underlying implicit measures also may provide deeper insights for the prediction of behavior by implicit measures.

REFERENCES

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "Would Jesse Jackson 'fail' the Implicit Association Test?" *Psychological Inquiry*, *15*, 257–278.

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: the case of shy behavior. *Journal of Personality and Social Psychology*, *83*, 380.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*, 136–141.

Banse, R., Gawronski, B., Rebetez, C., Gutt, H., & Morton, J. B. (2010). The development of spontaneous gender stereotyping in childhood: relations to stereotype knowledge and stereotype flexibility. *Developmental Science*, *13*, 298–306.

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie*, *48*, 145–160.

Bar-Anan, Y., & Nosek, B. A. (2012). Reporting intentional rating of the primes predicts priming effects in the Affective Misattribution Procedure. *Personality and Social Psychology Bulletin*, *38*, 1194–1208.

Bar-Anan, Y., Nosek, B. A., & Vianello, M. (2009). The sorting paired features task: a measure of association strengths. *Experimental Psychology*, *56*, 329–343.

Bargh, J. A. (1994). The four horsemen of automaticity: awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 1–40). Hillsdale, NJ: Erlbaum.

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *Psychological Record*, *60*, 527–542.

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, *17*, 53–58.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*, 242–261.

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: the moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*, 828–841.

Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: the blind men and the elephant revisited? *Journal of Personality and Social Psychology*, *79*, 631–643.

Brendl, C. M., Markman, A. B., & Messner, C. (2005). Indirectly measuring evaluations of several attitude objects in relation to a neutral

reference point. *Journal of Experimental Social Psychology*, *41*, 346–368.

Briñol, P., Petty, R. E., & Wheeler, S. C. (2006). Discrepancies between explicit and implicit self-concepts: Consequences for information processing. *Journal of Personality and Social Psychology*, *91*, 154–170.

Brochu, P. M., Gawronski, B., & Esses, V. M. (2011). The integrative prejudice framework and different forms of weight prejudice: an analysis and expansion. *Group Processes and Intergroup Relations*, *14*, 429–444.

Calanchini, J., Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2013). Counterprejudicial training reduces activation of biased associations and enhances response monitoring. *European Journal of Social Psychology*, *43*, 321–325.

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: a meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review*, *16*, 330–350.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, *25*, 215–224.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428.

Cone, J., & Ferguson, M. J. (2015). He did *what*? The role of diagnosticity in revising implicit evaluations. *Journal of Personality and Social Psychology*, *108*, 37–57.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating multiple processes in implicit social cognition: the Quad-Model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469–487.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349.

Cunningham, W. A., Nezlek, J. B., & Banaji, M. R. (2004). Implicit and explicit ethnocentrism: revisiting the ideologies of prejudice. *Personality and Social Psychology Bulletin*, *30*, 1332–1346.

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: a multilevel framework for attitudes and evaluation. *Social Cognition*, *25*, 736–760.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–814.

De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology*, *50*, 77–85.

De Houwer, J. (2006). Using the Implicit Association Test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Learning and Motivation*, *37*, 176–187.

De Houwer, J. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass*, *8*, 342–353.

De Houwer, J., & De Bruycker, E. (2007a). The identification-EAST as a valid measure of implicit attitudes toward alcohol-related stimuli. *Journal of Behavior Therapy and Experimental Psychiatry*, *38*, 133–143.

De Houwer, J., & De Bruycker, E. (2007b). The Implicit Association Test outperforms the Extrinsic Affective Simon Task as an implicit measure of interindividual differences in attitudes. *British Journal of Social Psychology*, *46*, 401–421.

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology*, *24*, 252–287.

De Houwer, J., Heider, N., Spruyt, A., Roets, A., & Hughes, S. (2015). The relational responding task: toward a new implicit measure of beliefs. *Frontiers in Psychology*, *6*, 319.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: a normative analysis and review. *Psychological Bulletin*, *135*, 347–368.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: a review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*, 853–869.

Degner, J. (2009). On the (un)controllability of affective priming: strategic manipulation is feasible but can possibly be prevented. *Cognition and Emotion*, *23*, 327–354.

Degner, J., & Wentura, D. (2008). The extrinsic affective Simon task as an instrument for indirect assessment of prejudice. *European Journal of Social Psychology*, *38*, 1033–1043.

De Raedt, R., Schacht, R., Franck, E., & De Houwer, J. (2006). Self-esteem and depression revisited: implicit positive self-esteem in depressed patients? *Behaviour Research and Therapy*, *44*, 1017–1028.

Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: negation as reflective operation. *Journal of Personality and Social Psychology*, *91*, 385–405.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*, 62–68.

Eder, A. B., & Rothermund, K. (2008). When do motor behaviors (mis)match affective stimuli? An evaluative coding view of approach and avoidance reactions. *Journal of Experimental Psychology: General*, *137*, 262.

Egloff, B., & Schmukle, S. C. (2002). Predictive validity of an Implicit Association Test for assessing anxiety. *Journal of Personality and Social Psychology*, *83*, 1441–1455.

Egloff, B., & Schmukle, S. C. (2003). Does social desirability moderate the relationship between implicit and explicit anxiety measures? *Personality and Individual Differences*, *35*, 1697–1706.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, *25*, 603–637.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013–1027.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 229–238.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, *58*, 203–210.

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology*, *27*, 307–316.

Friese, M., Hofmann, W., & Schmitt, M. (2008). When and why do implicit measures predict behavior? Empirical evidence for the moderating role of opportunity, motivation, and process reliance. *European Review of Social Psychology*, *19*, 285–338.

Galdi, S., Gawronski, B., Arcuri, L., & Friese, M. (2012). Selective exposure in decided and undecided individuals: differential relations to automatic associations and conscious beliefs. *Personality and Social Psychology Bulletin*, *38*, 559–569.

Gast, A., & De Houwer, J. (2012). Evaluative conditioning without directly experienced pairings of the conditioned and the unconditioned stimuli. *Quarterly Journal of Experimental Psychology*, *65*, 1657–1674.

Gawronski, B. (2012). Back to the future of dissonance theory: cognitive consistency as a core motive. *Social Cognition*, *30*, 652–668.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: theory, evidence, and open questions. *Advances in Experimental Social Psychology*, *44*, 59–127.

Gawronski, B., Bodenhausen, G. V., & Becker, A. (2007). I like it, because I like myself: associative self-anchoring and post-decisional change of implicit evaluations. *Journal of Experimental Social Psychology*, *43*, 221–232.

Gawronski, B., Brannon, S. M., & Bodenhausen, G. V. (2017). The associative-propositional duality in the representation, formation, and expression of attitudes. In R. Deutsch, B. Gawronski, & W. Hofmann (Eds.), *Reflective and impulsive determinants of human behavior* (pp. 103–118). New York, NY: Psychology Press.

Gawronski, B., & Cesario, J. (2013). Of mice and men: what animal research can tell us about context effects on automatic responses in humans. *Personality and Social Psychology Review*, *17*, 187–215.

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 283–310). New York, NY: Cambridge University Press.

Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When "just say no" is not enough: affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology*, *44*, 370–377.

Gawronski, B., Geschke, D., & Banse, R. (2003). Implicit bias in impression formation: associations influence the construal of individuating information. *European Journal of Social Psychology*, *33*, 573–589.

Gawronski, B., Hofmann, W., & Wilbur, C.J. (2006). Are "implicit" attitudes unconscious? *Consciousness and Cognition*, *15*, 485–499.

Gawronski, B., Hu, X., Rydell, R. J., Vervliet, B., & De Houwer, J. (2015). Generalization and contextualization in automatic evaluation revisited: a meta-analysis of successful and failed replications. *Journal of Experimental Psychology: General*, *144*, e50–e64.

Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: when implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology*, *44*, 1355–1361.

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, *2*, 181–193.

Gawronski, B., & Payne, B. K. (Eds.). (2010). *Handbook of implicit social cognition: Measurement, theory, and applications*. New York, NY: Guilford Press.

Gawronski, B., Peters, K. R., Brochu, P. M., & Strack, F. (2008). Understanding the relations between different forms of racial prejudice: a cognitive consistency perspective. *Personality and Social Psychology Bulletin*, *34*, 648–665.

Gawronski, B., Peters, K. R., & LeBel, E. P. (2008). What makes mental associations personal or extrapersonal? Conceptual issues in the methodological debate about implicit attitude measures. *Social and Personality Psychology Compass*, *2*, 1002–1023.

Gawronski, B., Rydell, R. J., Vervliet, B., & De Houwer, J. (2010). Generalization versus contextualization in automatic evaluation. *Journal of Experimental Psychology: General*, *139*, 683–701.

Gawronski, B., & Sritharan, R. (2010). Formation, change, and contextualization of mental associations: determinants and principles of variations in implicit measures. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: measurement, theory, and applications* (pp. 216–240). New York, NY: Guilford Press.

Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: dissonance changes explicit, but not implicit, attitudes. *Journal of Experimental Social Psychology*, *40*, 535–542.

Gawronski, B., & Walther, E. (2008). The TAR effect: when the ones who dislike become the ones who are disliked. *Personality and Social Psychology Bulletin*, *34*, 1276–1289.

Gawronski, B., Walther, E., & Blank, H. (2005). Cognitive consistency and the formation of interpersonal attitudes: cognitive balance affects the encoding of social information. *Journal of Experimental Social Psychology*, *41*, 618–626.

Gawronski, B., & Ye, Y. (2015). Prevention of intention invention in the affect misattribution procedure. *Social Psychological and Personality Science*, *6*, 101–108.

Gawronski, B., Ye, Y., Rydell, R. J., & De Houwer, J. (2014). Formation, representation, and activation of contextualized attitudes. *Journal of Experimental Social Psychology*, *54*, 188–203.

Gemar, M. C., Segal, Z. V., Sagrati, S., & Kennedy, S. J. (2001). Mood-induced changes on the Implicit Association Test in recovered depressed patients. *Journal of Abnormal Psychology*, *110*, 282–289.

Gibson, B. (2008). Can evaluative conditioning change attitudes toward mature brands? New evidence from the Implicit Association Test. *Journal of Consumer Research*, *35*, 178–188.

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2009). Aging and prejudice: diminished regulation of automatic race bias among older adults. *Journal of Experimental Social Psychology*, *45*, 410–414.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, *108*, 553–561.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*, 3–25.

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*, 1022–1038.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41.

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, *90*, 1–20.

Grumm, M., Nestler, S., & von Collani, G. (2009). Changing explicit and implicit attitudes: the case of self-esteem. *Journal of Experimental Social Psychology*, *45*, 327–335.

Hahn, A., & Gawronski, B. (2014). Do implicit evaluations reflect unconscious attitudes? *Behavioral and Brain Sciences*, *37*, 28–29.

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General*, *143*, 1369.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: a meta-analysis. *Psychological Bulletin*, *136*, 390–421.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*, 1369–1385.

Hofmann, W., Gschwendner, T., Nosek, B. A., & Schmitt, M. (2005). What moderates implicit-explicit consistency? *European Review of Social Psychology*, *16*, 335–390.

Hofmann, W., Gschwendner, T., & Schmitt, M. (2005). On implicit-explicit consistency: the moderating role of individual differences in awareness and adjustment. *European Journal of Personality*, *19*, 25–49.

Hofmann, W., Gschwendner, T., Wiers, R., Friese, M., & Schmitt, M. (2008). Working memory capacity and self-regulation: towards an individual differences perspective on behavior determination by automatic versus controlled processes. *Journal of Personality and Social Psychology*, *95*, 962–977.

Hofmann, W., Rauch, W., & Gawronski, B. (2007). And deplete us not into temptation: automatic attitudes, dietary restraint, and self-regulatory resources as determinants of eating behavior.

Journal of Experimental Social Psychology, *43*, 497–504.

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorizing in implicit attitude research: propositional and behavioral alternatives. *Psychological Record*, *61*, 465–496.

Jordan, C. H., Spencer, S. J., Zanna, M. P., Hoshino-Browne, E., & Correll, J. (2003). Secure and defensive high self-esteem. *Journal of Personality and Social Psychology*, *85*, 969–978.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*, 774–788.

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, *91*, 16–32.

Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process-components of the Implicit Association Test: a diffusion model analysis. *Journal of Personality and Social Psychology*, *93*, 353–368.

Koole, S. L., Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a name: implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology*, *80*, 669–685.

Krieglmeyer, R., & Deutsch, R. (2010). Comparing measures of approach-avoidance behavior: the manikin task vs. two versions of the joystick task. *Cognition and Emotion*, *24*, 810–828.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberative judgments are based on common principles. *Psychological Review*, *118*, 97–109.

Maison, D., Greenwald, A. G., & Bruin, R. H. (2004). Predictive validity of the Implicit Association Test in studies of brands, consumer attitudes, and behavior. *Journal of Consumer Psychology*, *14*, 405–415.

Mann, T., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology*, *108*, 823–849.

McConahay, J. B. (1983). Modern racism and modern discrimination: the effects of race, racial attitudes, and context on simulated hiring decisions. *Personality and Social Psychology Bulletin*, *9*(4), 551–558.

Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the Implicit Association Test: the ReAL model for the IAT. *Journal of Personality and Social Psychology*, *104*, 45–69.

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, *132*, 455–469.

Moors, A., Spruyt, A., & De Houwer, H. (2010). In search for a measure that qualifies as implicit: recommendations based on a decompositional view of automaticity. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 19–37). New York, NY: Guilford Press.

Moran, T., & Bar-Anan, Y. (2013). The effect of object-valence relations on automatic evaluation. *Cognition and Emotion*, *27*, 743–752.

Moran, T., Bar-Anan, Y., & Nosek, B. A. (2015). Processing goals moderate the effect of co-occurrence on automatic evaluation. *Journal of Experimental Social Psychology*, *60*, 157–162.

Neumann, R., Hülsenbeck, K., & Seibt, B. (2004). Attitudes toward people with AIDS and avoidance behavior: automatic and reflective bases of behavior. *Journal of Experimental Social Psychology*, *40*, 543–550.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, *134*, 565–584.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition*, *19*, 625–666.

Nosek, B. A., & Hansen, J. J. (2008). The associations in our heads belong to us: searching for attitudes and knowledge in implicit evaluation. *Cognition & Emotion*, *22*, 553–594.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science*, *12*, 413–147.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: personalizing the IAT. *Journal of Personality and Social Psychology*, *86*, 653–667.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, *32*, 421–433.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: a meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*, 171–192.

Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology*, *20*, 272–314.

Payne, B. K., Brown-Iannuzzi, J., Burkley, M., Arbuckle, N. L., Cooley, E., Cameron, C. D., & Lundberg, K. B. (2013). Intention invention and the Affect Misattribution Procedure: reply to Bar-Anan and Nosek (2012). *Personality and Social Psychology Bulletin*, *39*, 375–386.

Payne, B. K., Burkley, M., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology*, *94*, 16–31.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. (2005). An inkblot for attitudes: affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277–293.

Payne, B. K., & Gawronski, B. (2010). A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going? In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: measurement, theory, and applications* (pp. 1–15). New York, NY: Guilford Press.

Payne, B. K., & Lundberg, K. B. (2014). The Affect Misattribution procedure: ten years of evidence on reliability, validity, and mechanisms. *Social and Personality Psychology Compass*, *8*, 672–686.

Penke, L., Eichstaedt, J., & Asendorpf, J. B. (2006). Single Attribute Implicit Association Tests (SA-IAT) for the assessment of unipolar constructs: the case of sociosexuality. *Experimental Psychology*, *53*, 283–291.

Perugini, M., Richetin, J., & Zogmeister, C. (2010). Prediction of behavior. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: measurement, theory, and applications* (pp. 255–277). New York, NY: Guilford Press.

Peters, K. R., & Gawronski, B. (2011a). Are we puppets on a string? Comparing the impact of contingency and validity on implicit and explicit evaluations. *Personality and Social Psychology Bulletin*, *37*, 557–569.

Peters, K. R., & Gawronski, B. (2011b). Mutual influences between the implicit and explicit self-concepts: the role of memory activation and motivated reasoning. *Journal of Experimental Social Psychology*, *47*, 436–442.

Petty, R. E., Briñol, P., & DeMarree, K. G. (2007). The meta-cognitive model (MCM) of attitudes: implications for attitude measurement, change, and strength. *Social Cognition*, *25*, 657–686.

Petty, R. E., Tormala, Z. L., Briñol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: an exploration of the PAST Model. *Journal of Personality and Social Psychology*, *90*, 21–41.

Ranganath, K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *Journal of Experimental Social Psychology*, *44*, 386–396.

Remue, J., Hughes, S., De Houwer, J., & De Raedt, R. (2014). To be or want to be: disentangling the role of actual versus ideal self in implicit self-esteem. *PLoS ONE*, *9*(9), e108837.

Richeson, J. A., & Ambady, N. (2003). Effects of situational power on automatic racial prejudice. *Journal of Experimental Social Psychology*, *39*, 177–183.

Richetin, J., Perugini, M., Adjali, I., & Hurling, R. (2007). The moderator role of intuitive versus deliberative decision making for the predictive validity of implicit and explicit measures. *European Journal of Personality*, *21*, 529–546.

Roefs, A., Quaedacjers, L., Werrij, M. Q., Wolters, G., Havermans, R., Nederkoorn, C., … Jansen, A. (2006). The environment influences whether high-fat foods are associated with palatable or with unhealthy. *Behaviour Research and Therapy*, *44*, 715–736.

Roth, J., & Steffens, M. C. (2014). When I becomes we: associative self-anchoring drives implicit intergroup bias in minimal groups. *Social Psychology*, *45*, 253–264.

Rothermund, K., Teige-Mocigemba, S., Gast, A., & Wentura, D. (2009). Minimizing the influence of recoding in the implicit association test: the recoding-free implicit association test (IAT-RF). *Quarterly Journal of Experimental Psychology*, *62*, 84–98.

Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science*, *13*, 79–82.

Rudman, L. A., Greenwald, A. G., Mellott, D. S., & Schwartz, J. L. (1999). Measuring the automatic components of prejudice: flexibility and generality of the Implicit Association Test. *Social Cognition*, *17*, 437–465.

Rudman, L. A., Phelan, J. E., & Heppen, J. B. (2007). Developmental sources of implicit attitudes. *Personality and Social Psychology Bulletin*, *33*, 1700–1713.

Rydell, R. J., & Gawronski, B. (2009). I like you, I like you not: understanding the formation of context-dependent automatic attitudes. *Cognition and Emotion*, *23*, 1118–1152.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: a systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*, 995–1008.

Rydell, R. J., & McConnell, A. R. (2010). Consistency and inconsistency in implicit social cognition: the case of implicit and explicit measures of attitudes. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: measurement, theory, and applications* (pp. 295–311). New York, NY: Guilford Press.

Rydell, R. J., McConnell, A. R., & Mackie, D. M. (2008). Consequences of discrepant explicit and implicit attitudes: cognitive dissonance and increased information processing. *Journal of Experimental Social Psychology*, *44*, 1526–1532.

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of two minds: forming and changing valence-inconsistent implicit and explicit attitudes. *Psychological Science*, *17*, 954–958.

Rydell, R. J., McConnell, A. R., Strain, L. M., Claypool, H. M., & Hugenberg, K. (2007). Implicit and explicit attitudes respond differently to increasing amounts of counterattitudinal information. *European Journal of Social Psychology*, *37*, 867–878.

Sava, F. A., Maricutoiu, L. P., Rusu, S., Macsinga, I., Virga, D., Cheng, C. M., & Payne, B. K. (2012). An inkblot for the implicit assessment of personality: the semantic misattribution procedure. *European Journal of Personality*, *26*, 613–628.

Schnabel, K., Banse, R., & Asendorpf, J. (2006). Employing automatic approach and avoidance tendencies for the assessment of implicit personality self-concept: the Implicit Association Procedure (IAP). *Experimental Psychology*, *53*, 69–76.

Schwarz, N. (2007). Attitude construction: evaluation in context. *Social Cognition*, *25*, 638–656.

Seibt, B., Häfner, M., & Deutsch, R. (2007). Prepared to eat: how immediate affective and motivational responses to food cues are influenced by food deprivation. *European Journal of Social Psychology*, *37*, 359–379.

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. A., & Groom, C. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, *115*, 314–335.

Sherman, J. W., Gawronski, B., & Trope, Y. (Eds.). (2014). *Dual-process theories of the social mind*. New York, NY: Guilford Press.

Sherman, J. W., Klauer, K. C., & Allen, T. J. (2010). Mathematical modeling of implicit social cognition: the machine in the ghost. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: measurement, theory,*

and applications (pp. 156–175). New York, NY: Guilford Press.

Smith, C. T., & Nosek, B. A. (2011). Affective focus increases the concordance between implicit and explicit attitudes. *Social Psychology*, *42*, 300–313.

Solarz, A. K. (1960). Latency of instrumental responses as a function of compatibility with the meaning of eliciting verbal signs. *Journal of Experimental Psychology*, *59*, 239–245.

Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology*, *56*, 283–294.

Stahl, C., & Degner, J. (2007). Assessing automatic activation of valence: a multinomial model of EAST performance. *Experimental Psychology*, *54*, 99–112.

Steffens, M. C. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology*, *51*, 165–179.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220–247.

Teige, S., Schnabel, K., Banse, R., & Asendorpf, J. B. (2004). Assessment of multiple implicit self-concept dimensions using the Extrinsic Affective Simon Task. *European Journal of Personality*, *18*, 495–520.

Teige-Mocigemba, S., & Klauer, K. C. (2008). "Automatic" evaluation? Strategic effects on affective priming. *Journal of Experimental Social Psychology*, *44*, 1414–1417.

Teige-Mocigemba, S., Klauer, K. C., & Rothermund, K. (2008). Minimizing method-specific variance in the IAT: a single block IAT. *European Journal of Psychological Assessment*, *24*, 237–245.

Teige-Mocigemba, S., Klauer, K. C., & Sherman, J. W. (2010). A practical guide to Implicit Association Tests and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: measurement, theory, and applications* (pp. 117–139). New York, NY: Guilford Press.

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: measurement, theory, and applications* (pp. 95–116). New York, NY: Guilford Press.

Whitfield, M., & Jordan, C. H. (2009). Mutual influences of explicit and implicit attitudes. *Journal of Experimental Social Psychology*, *45*, 748–759.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*, 101–126.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationships with questionnaire measures. *Journal of Personality and Social Psychology*, *72*, 262–274.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, *81*, 815–827.

Wojnowicz, M. T., Ferguson, M. J., Dale, R., & Spivey, M. J. (2009). The self-organization of explicit attitudes. *Psychological Science*, *20*, 1428–1435.

Zanon, R., De Houwer, J., & Gast, A. (2012). Context effects in evaluative conditioning of implicit evaluations. *Learning & Motivation*, *43*, 155–165.